Casual Indoor HDR Radiance Capture from Omnidirectional Images

Pulkit Gera[1]
pulkit.gera@research.iiit.ac.in
Mohammad Reza Karimi Dastjerdi[2]
mohammad.karimi-dastjerdi.1@ulaval.ca
Charles Renaud[2]
renaudcharles14@hotmail.com
P. J. Narayanan[1]
pjn@iiit.ac.in
Jean-François Lalonde[2]
jflalonde@gel.ulaval.ca

[1] International Institute of Information Technology
Hyderabad, India
[2] Université Laval
Quebec City, Canada

A Datasets & Training LDR2HDR

We train the LDR2HDR module on Laval Indoor Dataset. We augment the training set with random rotations (about the vertical axis), intensity changes (multiply the image by 2^{γ}, with $\gamma \sim \mathrm{U}(-0.1, 0.1)$) and exposure changes (make the median intensity of image $0.5+\gamma$). Here, $\mathrm{U}(a,b)$ indicates a uniform distribution in the $[a,b]$ interval. After augmentation, the resulting HDR panorama is used as target $\mathbf{t}$ for training. The input $\mathbf{x}$ is created by clipping $\mathbf{t}$ to the [0,1] interval. We further apply hue shift and unsharp mask (amount =1, $\sigma \sim \mathrm{U}(0,3)$), add small amount of per-pixel Gaussian noise ($\sigma = 0.01$), and augment the tonemapping $\mathbf{x}^{1+\gamma}$ to simulate the behaviour of a real camera. We train the network for 1500 epochs using the Adam [2] optimizer with a learning rate $\eta = 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 10^{-8}$.

Further, we finetune the model on a small dataset captured using Ricoh Theta Z to alleviate domain gap. We run 130 additional epochs using the Adam [2] optimizer with the same parameters as above.

B Training PanoHDR-NeRF

We use NeRF++ [3] as the basis of our project. An eight-layer MLP with 256 channels is used for predicting radiance and densities at the sampled points. Along each ray, we sample 64 points for training the coarse network and 128 points for training the fine network. The batch size of rays is 1024. We use integrated positional encoding to encode the inputs of the network as used in MipNeRF [1]. Similarly a single MLP is used to encode the scene. In addition, we also use spherical sampling, which weights pixels at the poles less with respect to pixels in the middle. The network is trained using the Adam optimizer [2] with learning

© 2022. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

rate $\eta = 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 10^{-8}$. The resolution of the training images is 960×480. The network is trained for approximately 500,000 iterations, which takes around 36 hours on a NVIDIA V100 GPU.

C Spherical Sampling

Equirectangular images correspond to the projection of a spherical signal onto a 2D plane, where (normalized) pixel coordinates (i, j) are related to azimuth $\varphi \in [-\pi, \pi]$ and elevation $\theta \in [-\pi/2, \pi/2]$ angles by

$$i = \frac{1}{2\pi}\varphi\cos\theta + \frac{1}{2}\,, \text{ and } j = \frac{\pi/2 - \theta}{\pi}\,. \tag{1}$$

To train PanoHDR-NeRF, we sample rays in spherical coordinates instead of pixel coordinates, where $\theta \sim \mathrm{U}(-\pi, \pi)$, $\varphi = \cos^{-1}(2\beta - 1)$, and $\beta \sim \mathrm{U}(0, 1)$.
We compare the results between planar and spherical sampling used for training PanoHDR-NeRF. We observe that spherical sampling performs much better and provides sharper results.

	Planar		Spherical	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Chess room	29.635	0.919	31.389	0.929
Stairway	25.381	0.892	27.381	0.891
Cafeteria	22.940	0.845	24.038	0.842
Spotlights	25.557	0.847	26.128	0.852
Dark class	29.955	0.911	31.368	0.917
Small class	29.231	0.906	30.611	0.911
Overall	27.115	0.886	28.486	0.902

Table 1: Quantitative comparison between planar and spherical sampling (on LDR images only) averaged over all captured scenes. Spherical sampling has better results.

D Video

We provide an additional video to showcase our results. We relight 3 virtual test objects made up of metal (armadillo), diffuse (bunny) and glass (sphere) with our recovered HDR images at novel viewpoints to demonstrate the dynamic range recovered. We present our results with a variety of scenes captured casually.

References

[1] Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *IEEE/CVF Int. Conf. Comput. Vis.*, 2021.

[2] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Int. Conf. Learn. Represent.*, 2015.

[3] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *CoRR*, abs/2010.07492, 2020.